

UNIVERSIDAD DEL VALLE DE GUATEMALA  
Facultad de Ingeniería



**Diseño e implementación de un repositorio de acceso público  
de datos y señales relacionados al estudio de la epilepsia**

Protocolo de trabajo de graduación presentado por Samuel Josue  
Silvestre Mendoza, estudiante de Ingeniería Mecatronica

Guatemala,

2022

## Resumen

Este protocolo presenta una propuesta de una nueva herramienta para el estudio de la epilepsia, continuando lo desarrollado en la fase II por Jorge Diego Manrique Sáenz, en la tesis “Herramienta de Software con una Base de Datos Integrada para el Estudio de la Epilepsia - Fase II”. Lo que se busca es llevar el concepto de la base de datos a un repositorio que se encuentre en la nube, y que sea de carácter público, de manera que contribuya a la comunidad científica en el estudio de la epilepsia y sus efectos.

A partir de la herramienta HUMANA y demás herramientas que nos permiten extraer datos para estudiar la epilepsia con métodos como electroencefalogramas, se buscará preparar el mejor espacio posible para que los usuarios puedan proveer o extraer datos de manera sencilla, y puedan contribuir a la comunidad de la manera que deseen. Esto se logrará por medio del cambio antes mencionado, de una base de datos local a un repositorio público, aumentando el alcance de la herramienta, permitiéndole a la comunidad ingresar y extraer datos con facilidad.

## Antecedentes

La epilepsia es una enfermedad que afecta al cerebro y las señales que envía, la perturbación en esta actividad neuronal causa en un aumento repentino de la actividad eléctrica que, puede llevar a causar convulsiones. Estas se alteran según el área del cerebro donde ha ocurrido [1].

A pesar de ser una de las condiciones neurológicas más comunes, más precisamente, llegando a afectar a más de 50 millones de personas alrededor del mundo (siendo más común en niños y mayores de 60 años) no se ha logrado comprender esta condición en su totalidad, teniendo como prueba de esto que, a día de hoy, no se posee un tratamiento que funcione para todos los casos. Existen más de 30 tipos de crisis epilépticas descritas con duración de unos segundos o minutos. Se pueden dividir en dos clases principales, convulsiones focales que afectan solamente una parte del cerebro, y convulsiones generalizadas que afectan todo el cerebro [2].

Cuando se habla de una base de datos relacional, se habla de que esta almacena y provee datos que están relacionados a otros. Se basan en un modelo fácil e intuitivo para acceder y almacenar la información en tablas. Cada fila es un registro con un numero de identificación único conocido como llave y posee un atributo con información específica al registro [3].

Hay dos clases de bases de datos en la nube, está la clase tradicional que es similar a una base de datos in-situ y administrada internamente, con una pequeña diferencia en el origen de la infraestructura, ya que la compañía compra un espacio virtual y un proveedor de servicios en la nube. La organización se encarga de manejar la base de datos. Luego está la base de datos como servicio (DBaaS), donde el proveedor ofrece más servicios de gestión de base de datos [4].

Actualmente, en terminos de procesos médicos, ya existen ejemplos de repositorios públicos de datos que ayudan a unir la comunidad en la investigación de distintas enfermedades y sus respectivos tratamientos. Tal es el caso del sitio llamado Physionet, que cuenta con

un gran catalogo de archivos de señales biológicas bien categorizadas, una gran colección de software de análisis de señales fisiológicas, además de tutoriales y material educativo [5].

El Centro de Epilepsia y Neurocirugía Funcional HUMANA, es una organización que se dedica a buscar el beneficio de sus pacientes que padecen problemas Neurológicos de difícil control tales como Epilepsia, Parkinson, Tumores Cerebrales, Columna Vertebral, Movimientos Anormales entre otros. Formada por profesionales de Neurociencias, HUMANA es el Centro de Referencia en Neurociencias para Guatemala y Centro América que posee los mejores recursos para el diagnóstico y tratamiento de enfermedades cerebrales [6].

Actualmente la universidad del Valle de Guatemala colabora con HUMANA en una estrecha relación que ha logrado que se consigan múltiples avances en áreas como la implementación de ingeniería biomédica en soluciones médicas.

Las tesis desarrollada por Jorge Diego Manrique [7] consistieron en mejorar lo realizado en la etapa uno desarrollada por María Fernanda Pineda [8] tratándose del desarrollo de una herramienta para análisis de señales electroencefalográficas y diseño de una base de datos para almacenar las señales grabadas por HUMANA. En la fase uno desarrollada por María Fernanda Pineda y María Jesús Aangulo, la base de datos por medio de 35 canales, con 20 obligatorios, almacenaba la información en tres tablas, una de datos del paciente, otra de datos descriptivos de la prueba, y otra con los datos de las señales en sus diferentes canales.

En la fase dos, esto fue expandido a poder almacenar características, clasificadores, anotaciones de las señales EEG, datos confidenciales, configuración para funcionalidad “Recuperar contraseña” y accesos de los usuarios. Además de esto, se realizaron múltiples mejoras a la seguridad, manejo de usuarios e interfaz de la aplicación en Matlab utilizada para obtener y guardar los datos obtenidos con HUMANA en la base de datos. Se implementó un diseño gráfico que permite entrar con distintos usuarios, crearlos y administrarlos, además de eliminarse la necesidad de entrar a través de un archivo .csv. También se crearon validaciones y programación defensiva para evitar errores en caso se ingresen datos no válidos o espacios en blanco. En la FIGURA 1 se puede ver el modelo relacional que utilizaron

Actualmente, la mayor dificultad se encuentra en proveer de forma sencilla la aplicación de Matlab, que es la que actualmente se utiliza, a todo el mundo. Se necesita conectar con un repositorio en la nube para evitar posibles pérdidas, y facilitar el acceso en cualquier parte, ya que actualmente se puede lograr el acceso remoto pero desde las computadoras que poseen HUMANA. Además, el tener un los datos solo de manera local imposibilita que todos los usuarios compartan la misma información. También se necesita optimizar el almacenamiento de archivos ya que por ejemplo, el archivo edf consume demasiado espacio.

## Justificación

En las etapas anteriores, sobre todo en la de Jorge Diego Manrique [7] se realizaron muchas mejoras en lo que respecta a la interfaz que permite el ingreso a la base de datos, tales como ampliar la cantidad de registros sobre el análisis, el paciente, las pruebas, el modelo utilizado, además de mejorar el ingreso de usuarios. Sin embargo, quedaron varias mejoras pendientes como mejorar la accesibilidad a la base de datos, optimizar la utilización de esta, además de expandir aún más su uso.

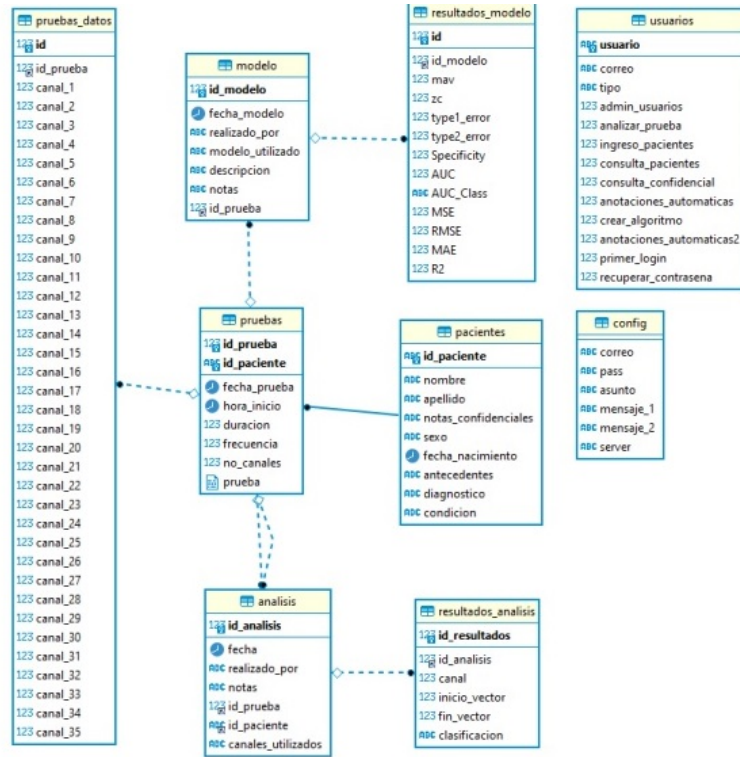


Figura 1: Modelo relacional utilizado en la fase anterior para la base de datos [7].

Actualmente la interfaz realizada por Jorge Diego Manrique está ajustada de buena manera a las herramientas de HUMANA, pero está pensada para un uso local, lo cual evita que se pueda compartir lo conseguido con la comunidad, además de limitar las posibilidades de esta aplicación.

Se busca crear un repositorio de señales y datos biomédicos relacionados a la epilepsia, de acceso público desde cualquier parte del mundo este debe estar bien organizado y con una interfaz sencilla para que cualquier persona pueda utilizar sus datos para sus propias investigaciones sobre la epilepsia.

El repositorio debe poder brindar a la comunidad aquello que ha investigado, y poder subir sus datos, con todos los datos que son necesarios tales como el método utilizado, el estado de lo tomado, el porcentaje de exactitud de los filtros en caso de tenerlo. De esta manera se busca aportar a la comunidad una herramienta que mejore el estudio de esta enfermedad de la que aún nos falta saber mucho.

## Objetivos

### Objetivo General

Desarrollar un repositorio de señales y datos biomédicos relacionados al estudio de la epilepsia, de acceso público en cualquier parte del mundo.

## Objetivos Específicos

- Evaluar modelos de repositorios biomédicos existentes y determinar el más adecuado para la implementación del repositorio propuesto.
- Obtener señales y datos biomédicos de pacientes con epilepsia de HUMANA, sin incluir información privada o que pueda identificar a los pacientes.
- Determinar e implementar el procesamiento y la organización de los datos y los formatos en que se almacenarán.
- Implementar el repositorio y las herramientas para su manejo, administración, y la interacción con este.

## Marco teórico

### Epilepsia

La epilepsia es una enfermedad que afecta al cerebro y las señales que envía. La perturbación en esta actividad neuronal causa en un aumento repentino de la actividad eléctrica que, puede causar convulsiones. Estas perturbaciones de la actividad neuronal varían según el área del cerebro donde ha ocurrido.

A pesar de ser una de las condiciones neurológicas más comunes, más precisamente, llegando a afectar a más de 50 millones de personas alrededor del mundo (siendo más común en niños y mayores de 60 años) no se ha logrado comprender esta condición en su totalidad, teniendo como prueba de esto que, a día de hoy, no se posee un tratamiento específico. De la población que sufre de epilepsia, un tercio sufren de epilepsia refractaria, lo cual se refiere a la clase de epilepsia en la que las convulsiones no pueden ser tratadas por los métodos antiepilépticos más utilizados o recomendados [1].

### Tipos de epilepsia

Existen más de 30 tipos de crisis epilépticas descritas con duración de unos segundos o minutos. Se pueden dividir en dos clases principales, convulsiones focales que afectan solamente una parte del cerebro, y convulsiones generalizadas que afectan todo el cerebro [2]. En la figura dos podemos ver un ejemplo de como se ve un encefalograma

En el caso de una crisis epiléptica focal o parcial, hay tres tipos: crisis parcial simple, compleja, y crisis parcial que desemboca en generalizada al extenderse a todo el cerebro. La crisis parcial simple es la que ocurre cuando se presenta alguna alteración de la memoria, el movimiento y las acciones. En este estado es probable que el afectado sepa lo que está pasando. La crisis parcial compleja es la más común, es en la que el afectado pierde el conocimiento, con repetición convulsiva de movimientos, es probable que el afectado no sepa que está sufriendo una convulsión. Normalmente, un médico puede interpretar que clase de epilepsia sufre el paciente a través de un electroencefalograma (EEG) y generar un plan de tratamiento [2].

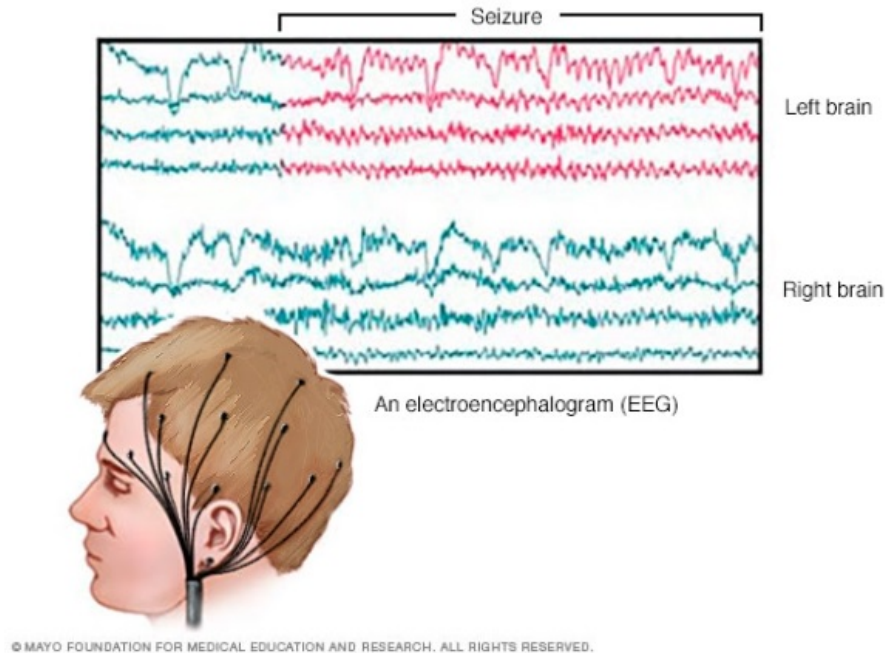


Figura 2: Actividad cerebral registrada por electroencefalograma [9].

## Electroencefalograma

Un electroencefalograma (EEG) es una prueba que detecta la actividad eléctrica del cerebro utilizando pequeños discos metálicos fijados en el cuero cabelludo. Las neuronas cerebrales se comunican a través de impulsos cerebrales y están activas en todo momento. Estas medidas se registran por medio de líneas onduladas en un reporte del encefalograma.

Un encefalograma puede ser utilizado para detectar cambios en la actividad cerebral, por medio de los cuales se pueden detectar enfermedades y trastornos cerebrales, especialmente la epilepsia u otros trastornos compulsivos. También puede detectar trastornos tales como daños cerebrales por lesiones en la cabeza, inflamación cerebral, trastorno de sueño, etc.

Un electroencefalograma también puede utilizarse para confirmar la muerte cerebral de una persona que se encuentra en un coma persistente. Además de esto, puede ser utilizado para encontrar la cantidad correcta de anestesia que se le debe dar a un paciente con coma inducido por medicamentos [9].

## Base de datos

Cuando se habla de una base de datos relacional, se habla de que esta almacena y provee datos que están relacionados a otros. Se basan en un modelo fácil e intuitivo para acceder y almacenar la información en tablas. Cada fila es un registro con un número de identificación único conocido como llave y posee un atributo con información específica al registro [3].

El objetivo de una base de datos es recolectar la información y mantenerla de tal forma

que pueda ser utilizada para análisis y operaciones futuras. Hay varios recursos que debe poseer como, por ejemplo, debe ser accesible para varios usuarios, la recolección debe ser conforme la información es generada, debe ser guardada la información aunque no sea usada en mucho tiempo, la lectura y escritura de información debe ser posible de hacer de forma constante, y la información debe estar relacionada para que sea fácil de buscar [3].

## Conceptos básicos

*Schemas:* Es una colección de temas en una base de datos, tiene un dueño y puede contener diversos objetos dentro de la base de datos como tablas, vistas, secuencias, etc [10].

*Tablas:* Unidad básica de almacenamiento en la base de datos, está formada por columnas con diferentes tipos de datos, y filas con la información a almacenar [11].

*Tipos de datos:* tipo de información que se quiere almacenar en la columna indicada en la tabla, por ejemplo: INT (número entero), Float (número de punto flotante), String (campo alfanumérico), etc. [12]

*Llaves primarias:* combinación de columnas que identifican cada registro de la tabla para que pueda ser modificado o consultado fácilmente por separado o en conjunto por otros registros [13].

*Llaves foráneas:* estas relacionan las tablas entre sí. La llave foránea es llamada “hija” y define las columnas que son relacionadas en la tabla “padre” [14].

## Base de datos en la nube

Una base de datos en la nube, es aquella que no se almacena en un equipo o sistema local, sino que se ejecuta desde la infraestructura de un proveedor de servicios.

Hay dos clases de bases de datos en la nube, está la clase tradicional que es similar a una base de datos in-situ y administrada internamente, con una pequeña diferencia en el origen de la infraestructura, ya que la compañía compra un espacio virtual de un proveedor de servicios en la nube. La organización se encarga de manejar la base de datos. Luego está la base de datos como servicio (DBaaS), donde el proveedor ofrece más servicios de gestión de base de datos [4].

## Ventajas

Escalables: el proveedor del servicio puede aumentar los recursos o proporcionar más espacio de almacenamiento si así lo desea el cliente [4].

Integridad de la información: Este tipo de almacenamiento de información puede replicar instancias de la base de datos, lo cual permite que varios usuarios al mismo tiempo puedan usarla, sin que los datos pierdan integridad. [4].

Ahorro de espacio físico: no es necesario instalar un equipo dedicado local, ya que todo

queda almacenado en los servidores del proveedor (de todas formas, tener un back-up local puede ser útil) [4].

Evitan problemas tales como la imposibilidad de acceder si se caen los servidores locales. Aumenta la seguridad: esto debido a que la información está protegida por empresas con reputación en el mundo digital [4].

Pero la ventaja más importante de todas es que se puede acceder a través de cualquier dispositivo desde cualquier parte del mundo [4].

## Metodología

- Investigar los distintos servicios para hostear el repositorio, y recopilar todas las características de cada servicio, en caso de no ser gratis, evaluar sus ventajas a comparación del precio
- Elegir el servicio basado en lo investigado, y prepararlo para que pueda ser utilizado por múltiples personas a la vez.
- Investigar el formato en que las herramientas de HUMANA y otras, graban los datos obtenidos de un encefalograma, y las diferencias entre un análisis con aprendizaje supervisado y uno no supervisado, además de los distintos métodos para estudiar la epilepsia.
- Preparar el servicio para que sea posible acceder desde cualquier parte del mundo.
- Planear la forma en que se van a organizar los datos.
- Preparar un repositorio prototipo en el cual se estudie la manera en que se le dará el acceso a cada usuario, de manera que sea lo más sencillo posible ingresar a los datos, o ingresar datos al repositorio.
- Seguir los pasos para unir nuestro repositorio a la biblioteca de repositorios de la universidad del valle de Guatemala.
- Estudiar la posibilidad de tener una interfaz grafica propia que se conecte a esa base de datos.

## Cronograma de actividades

La figura 3 muestra el cronograma de actividades. A continuación se describen las tareas a realizar.

### Descripción de tareas

#### *Tarea 1. Estudiar los prototipos de las fases anteriores*

Esta tarea trata de revisar y estudiar todo lo que se ha hecho con anterioridad en las etapas anteriores, para conocer qué se puede tomar de ellas.



	3 al 9 de Julio	10 al 16 de Julio	17 al 23 de Julio	24 al 30 de Julio	31 de Julio al 6 de agosto	7 al 13 de agosto	14 al 20 de agosto	21 al 27 de agosto	28 de agosto al 3 de septiembre	4 al 10 de septiembre	11 al 17 de septiembre	18 al 25 de septiembre
	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10	Semana 11	Semana 12
Tarea 1												
Tarea 2												
Tarea 3												
Tarea 4												
Tarea 5												
Tarea 6												
Tarea 7												
Tarea 8												
Tarea 9												
Tarea 10												

Figura 3: Cronograma de actividades

***Tarea 2. Estudiar los distintos modelos de repositorios y los servicios que los pueden albergar***

En esta tarea, se realiza un extenso estudio de distintos modelos de repositorios, incluyendo entre ellos los que utiliza la universidad del Valle de Guatemala, además de estudiar los distintos servicios para poder hospedar nuestro repositorio, se busca la mejor relación entre el acceso público, la seguridad del repositorio, y las herramientas que se puedan tener para implementarlo y permitir que los usuarios puedan acceder al repositorio y hace uso de sus datos, así como agregar nuevos.

***Tarea 3. Investigar los distintos formatos que se utilizan para los datos de electroencefalogramas y otros métodos del estudio de la epilepsia***

En esta tarea se hace una extensa investigación sobre la manera en que la epilepsia es estudiada, de manera de que se conozca el proceso en el que se obtienen los datos, el formato, como deben ser introducidos al repositorio para que sean legibles, que información de descripción suelen tener, etc.

***Tarea 4. Diseñar el formato en que será diseñado el repositorio***

Esta tarea se trata de tomar todo lo investigado anteriormente, y diseñar la manera en que los datos serán albergados, carpetas, tablas, la relación entre estas, etc. Básicamente es armar el modelo relacional del repositorio.

***Tarea 5. planear el acceso a la base de datos***

En esta etapa, lo que se busca es definir de que manera los usuarios entrarán a nuestra base de datos, de haber elegido un servicio que permita realizar una interfaz gráfica, ya sea por medio de una app o una página web, diseñar este recurso.

***Tarea 6. Prototipo 1***

Se implementa todo lo investigado anteriormente, de manera de que se pueda probar todo lo ideado. En este prototipo, lo que se busca es encontrar la mayor cantidad de posibles errores y problemas a futuro

***Tarea 7. Analisis prototipo 1***

Se realiza el análisis del prototipo antes mencionado, se buscan los errores y defectos del repositorio, se hace una lista de esto , y se idean maneras de resolver estos fallos.

***Tarea 8. Prototipos y análisis***

Se realiza un prototipo que resuelva estos fallos encontrados en la tarea anterior, esto se realiza en la primera semana de esta tarea. A su vez, se hace un análisis de este a la par, de encontrar errores, se modifica el prototipo. De esta manera, se va realizando prueba y error hasta que, todos los posibles problemas sean mitigados, y los errores corregidos.

***Tarea 9. Recomendaciones***

Se realizan las recomendaciones para una posible futura etapa. Se comenta que aspectos se pueden mejorar, que aspectos no fueron logrados, etc.

### ***Tarea 10, redacción del documento escrito***

El documento escrito se realiza durante todo el proceso.

## **Índice preliminar**

Prefacio

Lista de figuras

Lista de cuadros

Resumen

Abstract

1. Introducción

2. Antecedentes

3. Justificación

4. Objetivos

5. Alcance

6. Marco teórico

a) Epilepsia

b) Tipos de epilepsia

c) Tipos de convulsiones

d) Electroencefalograma

e) Base de datos

f) Organización de una base de datos

g) Conceptos básicos

h) servicios en la nube

i) servicios de repositorios de forma publica, como gestionarlos

j) seguridad en bases de datos publicas

k) posibilidades en el servicio de la nube escogido

7. diseño de prototipos

8. pruebas

a) aplicación de la oprganización de los datos

- b) poner en marcha la base de datos
- c) probar distintos accesos desde distintas partes
- d) poner a prueba la capacidad de ingreso de usuarios al mismo tiempo
- e) estudiar que todas las características del servicio funcionen

9.Resultados

10. Conclusiones

11. Recomendaciones

12. Bibliografía

13. Anexos

14. Glosario

## Referencias

- [1] C. E. Stafstrom, “Seizures and epilepsy: an overview for neuroscientists,” *PubMed*, 2015.
- [2] BUPA, “Epilepsia,” *BUPA Global Latinoamerica*, 2021.
- [3] Oracle, “What is a Relational Database?” <https://www.oracle.com/database/what-is-a-relational-database/>, 2022.
- [4] —, “Base de datos en la nube,” <https://www.oracle.com/es/database/what-is-a-cloud-database/>, 2022.
- [5] Physionet, “<https://physionet.org/about/>,” 2022.
- [6] HUMANA, “Especialistas en enfermedades neurológicas de difícil control,” *humanagt.org.*, 2022.
- [7] J. Manrique, “Herramienta de Software con una Base de Datos Integrada para el Estudio de la Epilepsia - Fase II,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2021.
- [8] M. Pineda, “Diseño e Implementación de una Base de Datos de Señales Biomédicas de Pacientes con Epilepsia,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2021.
- [9] M. Clinic, “EEG (electroencefalograma),” <https://www.mayoclinic.org/es-es/tests-procedures/eeg/about/20393875/>, 2022.
- [10] Oracle, “Introduction to Schema Objects,” <https://docs.oracle.com/cd/B1930601/server.102/b14220/>, 2022.
- [11] —, “Managing Tables,” <https://docs.oracle.com/cd/B1930601/server.102/b14231/tables.htm>, 2022.
- [12] —, “Selecting a Datatype,” <https://docs.oracle.com/cd/A5861701/server.804/a58241/ch5.htm>, 2022.

- [13] —, “Primary Keys,” *<https://docs.oracle.com/en/database/other-databases/nosql-database/12.2.4.5/driver-table/primary-keys.html>**GUID-6A063474-4A3A-4981-B1D8-71D0D95BE8DF*, 2022.
- [14] —, “*[https://docs.oracle.com/cd/E17952\\_01/mysql-5.6-en/create-table-foreign-keys.html](https://docs.oracle.com/cd/E17952_01/mysql-5.6-en/create-table-foreign-keys.html)*,” 2022.